

The OCDF diagram. A metamodel for object-oriented systems visual design.

Alexander Reshytko

IBM Science and Technology Center, Moscow

areshytko@ru.ibm.com

Abstract

We present a metamodel for modeling control and data flows on subclass scales in object-oriented systems. UML Profiles were used as a representation mean and a complete metamodel definition was provided with an example of a diagram application.

1 Introduction

One can hardly deny the necessity of modeling in software engineering in the majority of product life-cycle phases along with related tasks that occur in this field - from designing a new system, to refactoring or even reverse engineering an existing one. The key concept of every model type is abstraction. We describe the system of interest in a very particular context from a very specific angle hiding all other aspects. Models thus make important implicit information evident, explicit. And more models we have for a system - more important characteristics of it are clear and visible to us. Another important thing to note here is that each and every model type is a language. And like every language it determines the way we think about the system. It determines the way we design them, the way we discuss them and the way we try to understand them.

This article deals with the modeling of object-oriented systems. In this field UML plays an important role for over a decade now and it is highly promoted by OMG. It is an ISO standard and it has an abundant amount of literature about it as well

as tools to support it. Probably the main reason for such monopoly is the fact that a wide variety of object-oriented systems aspects can be modeled by UML. We simply don't need anything else in most of the cases. However, there are limitations and certain aspects of object-oriented systems are only vaguely covered by traditional UML toolset. To address this, the new metamodel was developed - OCDF (object control-data-flow).

Let's begin by recapping some ideas behind object-oriented systems and their constituents.

1.1 Object-oriented systems

One of the common mathematical formalizations of an object-oriented methodology is via a notion of an Abstract Data Type (ADT)[6]. This concept is pretty self-explanatory. It is a data type hence it's incorporated in a programming language type system. And it is defined implicitly in a representation free form. It is achieved via an ADT specification:

- A type name (Types section);
- operations applicable to instances of the type given solely by their signature (Functions section);
- predicates which define function invariants (Axioms section);
- characteristic functions which define domains of partial functions (in case some of the functions listed in the section 2 are partial) (Preconditions section).

Now, a class is merely a tuple of three things: an ADT specification, a representation choice and a mapping from functions listed in the specification to the representation in the form of a set of mechanisms, or features, each implementing one of the functions in terms of the representation, so as to satisfy the axioms and preconditions. And an object-oriented software construction is the building of software systems as structured collections of possibly partial abstract data type implementations.[6]

Class representation features can be divided into two separate kinds:

- those, that are represented by space, that is to say by associating a certain piece of information with every instance of the class [6]. They are usually called attributes;
- those, that are represented by time, that is to say by defining a certain computation (an algorithm) applicable to all instances of the class [6]. They are usually called routines or methods.

This distinction is very important and concerns not only computation time / memory consumption dilemma. Attributes form the values space of class instances. They constitute the possibility for distinction between class instances. They are the reason why we talk about classes and their instances and not just collections of functions amended to a programming language's type system. Moreover, in the imperative paradigm attributes form a state of a class instance. This is how an instance can be changed during its lifetime in most of the imperative object-oriented languages nowadays.

Now, if we think about kinds of relations that may occur between class representation features two most obvious candidates that come into mind are:

- control-flow relationship between routines - caller-callee relationship;
- data-flow relationship between routines or an attribute and a routine - supplier-provider relationship of a data in a form of function actual arguments, feature call result or a data value reading.

1.2 Motivation

We state that there are common use cases and situations when sub class structures formed by elements described above when modeled in a visual format can bring a lot of insights about the system being modeled. This is the idea behind OCDF models.

Ideally for the good granulated design with meaningful naming conventions the static UML Class diagram is able to provide a good understanding of the system. Nevertheless, classes with a rather complex logic or numerous responsibilities (the God object antipattern is an extreme case) are a fairly common phenomenon nowadays. This is especially applicable to the long living legacy code. In other words, there is a practical task to support, refactor, fix, understand (and share this understanding) them. And metamodels that deal with inter-class relationship very often provide a very little help here.

One of the dangerous drawbacks of the locking-based threading model which is the main multi-threading facility for many software engineering systems is a race condition possibility. And existing UML metamodels provide no facilities to cope with it.

Depicting loosely and strongly interconnected class substructures also helps in refactoring tasks and we found that there's often a functional reason behind such substructures which can be reflected in a more refined class system as a result of a refactoring process.

2 OCDF diagram

OCDF can help in such use cases and situations. In this metamodel a class is drawn as a directed graph of its features (nodes) interconnected with control and data flows (arcs). It is a static diagram that is designed to accomplish legacy code understanding and support and refactoring tasks. It models the implementation level of a system.

What makes OCDF useful is its ability to explicitly show tightly bound substructures of a class and boundaries between them. These are the boundaries that can be used to split the class into multiple ones during the refactoring process. These interconnected

substructures fairly often embody different functional responsibilities and thus helps a reader (and an author) revealing such information.

Moreover we state that not only the model itself provides a good understanding of a class but the way the diagram is filled supports and structures it. It provides a gradual way of understanding a class inner logic producing artifacts incrementally as this understanding increases.

3 The format of the metamodel

We considered several meta-metamodeling frameworks as a tool to describe the OCDF metamodel: MOF, EMF, UML Profiles, and have chosen a latter one. This decision was determined by two factors: a simplicity of definition and an available tools support.

The formal specification of the OCDF metamodel defined as the UML Profile will now be given.

4 The OCDF UML profile specification

4.1 Overview

This profile describes the Object Control Data Flow (OCDF) metamodel for object-oriented systems. The profile is aimed to combine representational features of UML with Control Flow Diagram (CDF) and Data Flow Diagram (DFD).

A DFD is a graphical representation of the "flow" of data through an information system, modeling its process aspects. It can be viewed as a directed graph wherein the nodes are external entities, processes or data stores and the edges are data flows [3].

Control flow differs from the Control Flow Graph and in a more high-level precision is just a caller-callee relationship between methods

The general idea behind OCDF models is to depict data and control flows in object-oriented systems on a sub-class scale. There are five basic metamodel elements to accomplish that:

- Class Data member - represents the traditional OOP concept of a basic element of an object's

state. It should have a one-to-one correspondence to modeled class's data members

- Class method - instances of this element in a model should have a one-to-one correspondence to modeled class's methods
- Data flow link - represents a directed data flow relationship between either a data member and a method or two methods.
- Control flow link - represents a directed control flow relationship between two methods.
- Class itself - basically a container for instances of all previously defined elements

Because of the nature of a UML profile these elements are defined as UML stereotypes which extend specific UML Superstructure metamodel elements. So they reuse and extend the standard UML semantics.

Although the most common use case is to describe a single class in the such type of diagrams there aren't any formal constraints and multiple classes can be modeled in a single diagram.

4.2 Profile diagram

Further are logical and formal diagrams for the metamodel described above. See Fig. 1 and Fig. 2

4.3 Stereotype Descriptions

4.3.1 ControlFlow

Generalizations: None.

Description: A ControlFlow is a special kind of a dependency relationship that may occur between instances of a MethodClass and signifies a caller-callee relationship.

Attributes: None.

Extensions: Dependency (from Dependencies)

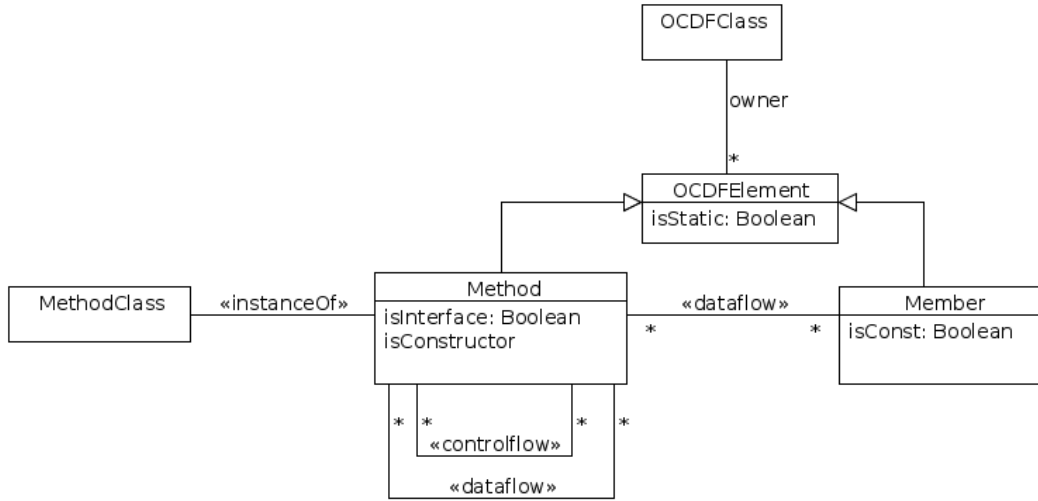


Figure 1: Informal structure

Constraints: [1] This relationship should occur only between method instances:

```

self.client->forAll(getAppliedStereotypes()-
>forAll(superClass->exists(name = 'ocdfmethod'))))
and self.supplier->forAll(getAppliedStereotypes()-
>forAll(superClass->exists(name =
'ocdfmethod'))))

```

Semantics: The semantic of a control flow relationship is dependent on a programming language that implements a system being modeled and should mean a method call.

Notation: A control flow is shown as a dashed arrow between two method instances. The method instance at the tail of the arrow (caller) calls the method instance at the arrowhead (the callee). The arrow may be labeled with an optional name (as any dependency relation).

4.3.2 DataFlow

Generalizations: None.

Description: A DataFlow is a kind of Information-Flow relationship that represents a directed data flow between either a data member and a method or two methods.

Attributes: None.

Extensions: InformationFlow (from Information-Flows)

Constraints: [1] Only constructors are allowed to modify constant data members

```

self.informationTarget->forAll(t |
t.getAppliedStereotype('ocdf::member') <> null
implies
t.getValue(t.getAppliedStereotype('ocdf::member'),
'isConst') = false or
self.informationSource->forAll(s |
s.getValue(s.getAppliedStereotype('ocdf::method'),
'isConstructor') = true or
s.getValue(s.getAppliedStereotype('ocdf::interfacemethod'),
'isConstructor') = true))

```

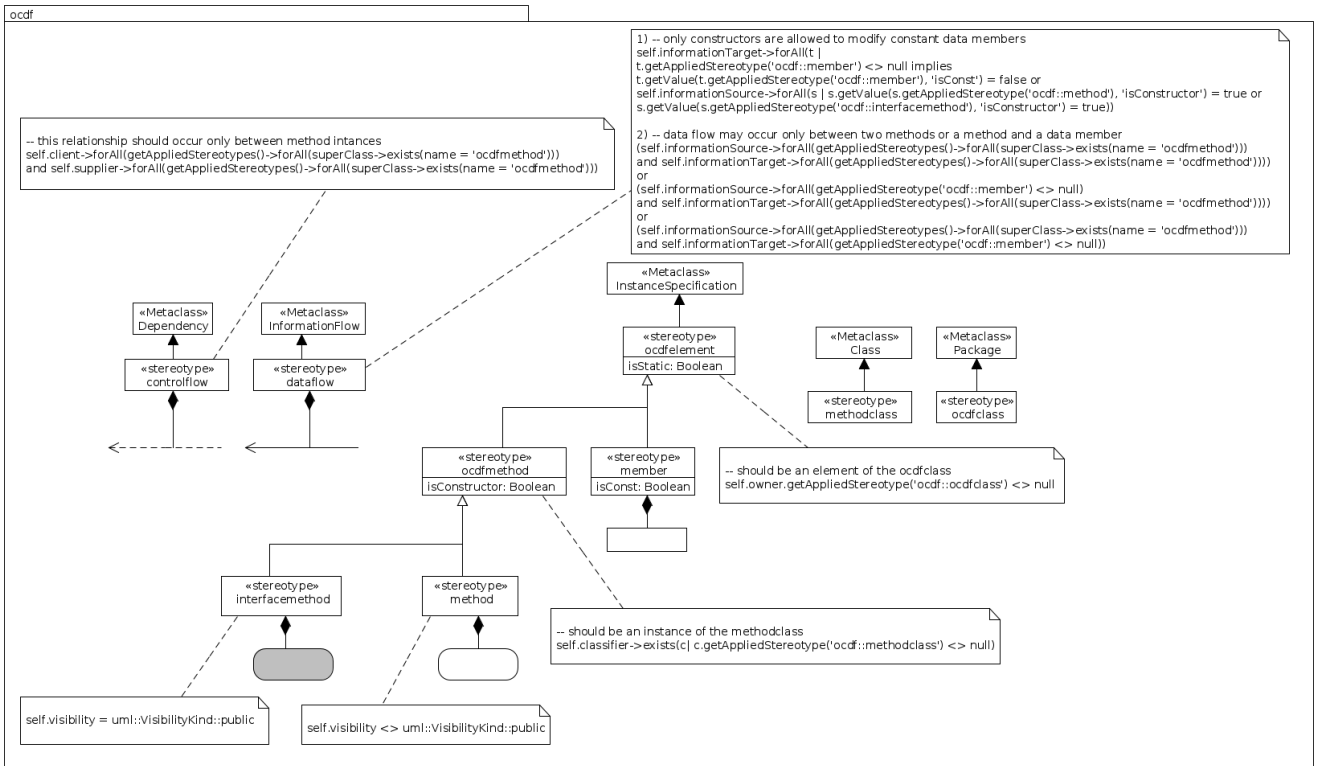


Figure 2: OCDF profile diagram

[2] Data flow may occur only between two methods or a method and a data member

```
(self.informationSource-
>forAll(getAppliedStereotypes()-
>forAll(superClass->exists(name = 'ocdfmethod'))))
and self.informationTarget-
>forAll(getAppliedStereotypes()-
>forAll(superClass->exists(name =
'ocdfmethod')))) or (self.informationSource-
>forAll(getAppliedStereotype('ocdf::member') <>
null) and self.informationTarget-
>forAll(getAppliedStereotypes()-
>forAll(superClass->exists(name =
'ocdfmethod')))) or (self.informationSource-
>forAll(getAppliedStereotypes()-
>forAll(superClass->exists(name = 'ocdfmethod'))))
and self.informationTarget-
>forAll(getAppliedStereotype('ocdf::member') <>
null))
```

Semantics: The semantics is dependent on a programming language that implements a system being modeled and should mean either of:

- a data read of a data member by a method;
- a result of a method call;
- a data being passed as an argument to a method call.

Notation: A data flow is shown as an arrow between two OCDF elements. The element at the tail of the arrow signifies a data provider and the element at the arrowhead signifies a data consumer.

4.3.3 InterfaceMethod

Generalizations: OCDFMethod

Description: Represents an interface method of a modeled class.

Attributes: None.

Extensions: InstanceSpecification (from Kernel)

Constraints: [1] Should have a public visibility
self.visibility = uml::VisibilityKind::public

Semantics: Should have a one-to-one correspondence to a modeled class's non-interface methods.

Notation: Is shown as a gray coloured rectangle with rounded corners, with the method signature shown inside the rectangle.

4.3.4 Member

Generalizations: OCDFElement

Description: Represents a data feature (attribute) of a modeled class.

Attributes: isConst: Boolean

Constness of a data member means that it can't be changed once it's value is defined in a constructor. But the semantic of this flag is generally dependent on a programming language that implements a modeled system

Extensions: InstanceSpecification (from Kernel)

Constraints: No additional constraints.

Semantics: Should have a one-to-one correspondence to modeled class's attributes.

Notation: Class data attributes are modeled as usual UML class instances - in a rectangular box with a name, a visibility prefix and a type inside.

4.3.5 Method

Generalizations: OCDFMethod

Description: Represents a non-interface method of a modeled class.

Attributes: None.

Extensions: InstanceSpecification (from Kernel)

Constraints: [1] Should have a non-public visibility

`self.visibility <> uml::VisibilityKind::public`

Semantics: Should have a one-to-one correspondence to modeled class's non-interface methods.

Notation: Is shown as a rectangle with rounded corners, with the method signature shown inside the rectangle.

4.3.6 MethodClass

Generalizations: None.

Description: Represents a class of callable methods.

Attributes: None.

Extensions: Class (from Kernel)

Constraints: No additional constraints.

Semantics: Instances of this element in a model should have a one-to-one correspondence to a modeled class's methods.

4.3.7 OCDFClass

Generalizations: None.

Description: Represents a class being modeled.

Attributes: None.

Extensions: Package (from Kernel)

Constraints: No additional constraints.

Semantics: Should have a one-to-one correspondence to a class in some object-oriented language.

4.3.8 OCDFElement

Generalizations: None.

Description: Represents an element of the OCDFClass. Abstract.

Attributes: isStatic: Boolean

In some object-oriented languages there is a notion of a static class feature - a feature of a class itself and not of its instances. This boolean attribute specifies this option.

Extensions: InstanceSpecification (from Kernel)

Constraints: [1] Should be an element of the ocdfclass.

`self.owner.getAppliedStereotype('ocdf::ocdfclass')`
`<> null`

Semantics: This is an abstract metamodel element that represents a modeled class's features.

4.3.9 OCDFMethod

Generalizations: OCDFElement

Description: Represents a callable feature (routine, method) of a modeled class. Abstract.

Attributes: isConstructor: Boolean

The notion of a constructor in many object-oriented languages means a method used to create an instance of a class and usually has a separate call syntax.

Extensions: InstanceSpecification (from Kernel)

Constraints: [1] Should be an instance of the MethodClass.

`self.classifier->exists(c|`
`c.getAppliedStereotype('ocdf::methodclass') <>`
`null)`

4.4 Usage recommendations and suggestions

As was already said the main purpose of this metamodel is to provide a help in understanding the legacy code, i.e. reverse-engineer it. And so the diagram should accompany its author in the process of investigating a class. Here a list of possible recommendations is provided:

- The diagram has shown a good performance in combination with the UML Communication Diagram - for depicting both inter-class and in-class control flows.
- Modeling can be conveyed on different abstraction levels:
 - just data flows between data members and methods
 - data flow involving data members and control flow between methods
 - the most detailed way involving all data flows between methods also together with control flows (separate flow arrows for input parameters and output results) allowing for code generation and or obtained automatically from the source code (depicting all this information by hand might be tedious)
- Inheritance requires a special treatment. We have found that a most convenient way is to depict parent features "lazily" by necessity - only if it's used in a current modeled child class explicitly.

5 An example of OCDF diagram

As an example of an OCDF diagram application we have chosen the FileManager class from the Clang project [2]. The Subversion revision number is 212447. See Fig. 3

6 Conclusions

We introduced a new metamodel for modeling control and data flows on subclass scales in object-oriented systems. The UML Profiles were used to define the metamodel which means that OCDF diagram can be used by any tool that conforms UML 2 specifications.

References

- [1] OMG Unified Modeling Language™, Superstructure. Version 2.4.1. Object Management Group, Inc., August 2011., <http://www.omg.org/spec/UML/2.4.1/Superstructure/PDF>, accessed: Dec. 2014.
- [2] Clang, <http://clang.llvm.org/>, accessed: Dec. 2014.
- [3] Bruza, P. D., Van der Weide, Th. P., "The Semantics of Data Flow Diagrams", University of Nijmegen, 1993.
- [4] E.D. Falkenberg, R. van der Pols, and Th. P. van der Weide, "Understanding process structure diagrams. Information Systems", 16(4):417-428, Sept. 1991.
- [5] Frances E. Allen, "Control flow analysis". SIGPLAN Notices 5 (7), doi:10.1145/390013.808479, Jul. 1970, pp. 1-19
- [6] Meyer, Bertrand, "Object-Oriented Software Construction, second edition", Prentice Hall. ISBN 0-13-629155-4, 1997
- [7] Deborah J. Armstrong, "The Quarks of Object-Oriented Development.", Communications of the ACM, 49(2):123-128, February 2006.

